2Л

УДК 519.6

О.Н. Литвин, Е.В. Ярмош
Украинская инженерно-педагогическая академия, г. Харьков
Украина, 61003, г. Харьков, ул. Университетская 16

Оценка ошибки округления приближения функций двух переменных одномерными операторами

O.M. Lytvyn, O.V. Iarmosh
Ukrainian Engineering Pedagogics Academy, c. Kharkiv
Ukraine, 61003, c. Kharkiv, Universitetska st., 16

Estimation of Rounding Errors of Two Variables Functions Approximation of One-Dimensional Operators

О.М. Литвин, О.В. Ярмош
Українська інженерно-педагогічна академія, м. Харків
Україна, 61003, м. Харків, вул. Університетська, 16

Оцінка похибки заокруглення наближення функцій двох змінних одномірними операторами

В статье предложен подход к оценке ошибки округления, возникающей при приближении функции двух переменных одномерными операторами смешанной аппроксимации при решении системы линейных алгебраических уравнений с помощью обратной матрицы.
Ключевые слова: одномерный оператор, смешанная аппроксимация, ошибка округления, ошибка метода, ошибка данных, обратная матрица.

This paper proposes an approach to the assessment of rounding errors arising in the approximation of a function of two variables-dimensional operators, by solving a system of linear algebraic equations, which can be reduced to the computation of the inverse matrix.
Key words: one-dimensional operator, blending approximation, rounding error, error of the method, data error, inverse matrix

В статті запропоновано підхід до оцінки похибки заокруглення, що виникає при наближенні функції двох змінних одновимірними операторами, шляхом розв’язання системи лінійних алгебраїчних рівнянь, що зводиться до обчислення оберненої матриці.
Ключові слова: одновимірний оператор, змішана апроксимація, похибка заокруглення, похибка методу, похибка даних, обернена матриця

Введение

С каждым годом сложность проведения вычислительных расчетов на ЭВМ увеличивается, возникает необходимость решения научных и прикладных задач, связанных с обработкой больших массивов входных данных. Все более актуальной становится проблема оценки ошибки округления при приближении функции двух и более переменных одномерными операторами, возникающей при расчетах на ЭВМ, и определение ее влияния на общий результат с учетом ошибок метода вычисления и входящих данных.

Сегодня широко используются интерполяционные и аппроксимационные методы приближения функций, среди которых можно отметить приближение функций двух переменных, заданных следами на системе взаимно перпендикулярных прямых или набором дискретных данных через операторы, действующие на функцию по одной переменной [1]. В дальнейшем будем называть такие операторы одномерными.

Цель работы – разработать метод оценки ошибки округления приближения функции двух переменных заданных следами на системе взаимно перпендикулярных прямых, через одномерные операторы, рассмотренные в работе [1], учитывая вид операторов и предположение, что соответствующие интегралы вычислены точно. Считаем, что ошибка округления возникает в ходе решения систем $B_1 u = F_1^T(y)$ и $vB_2 = F_2(x)$ путем нахождения решения с помощью обратных матриц.

Постановка задачи

В работе [2] авторы доказали теорему об ошибке приближения, включая ошибку метода приближения и ошибку входящих данных, функции двух переменных $f(x,y) \in C^{2,2}[0,1]^2$ оператором сплайн-интерлинации вида

$$\widetilde{O}f(x,y) = \sum_{k=0}^{N} \varphi_k(y) h_{1,k}(x) + \sum_{l=0}^{N} \psi_l(x) h_{2,l}(y) - \sum_{k=0}^{N}\sum_{l=0}^{N} \widetilde{f}_{k,l} h_{1,k}(x) h_{2,l}(y), \qquad (1)$$

когда следы $f(x_k, y)$, $f(x, y_l)$ заданы функциями $\varphi_k(y)$, $\psi_l(x)$ с ошибками и значения $f(x_k, y_l)$ также заданы числами $\widetilde{f}_{k,l}$ с ошибками. Рассмотрим подходы к оценке ошибки округления при применении разработанного в [2] метода приближения функций двух переменных, заданных следами на системе взаимно перпендикулярных прямых, когда оператор (1) задан в виде

$$Z^* f(x,y) = A_1 f(x,y) + A_2 f(x,y) - A_1 A_2 f(x,y), \qquad (2)$$

где одномерные операторы A_1, A_2 задают наилучшее среднеквадратическое приближение функции $f(x,y)$ по переменным x и y, и определяются формулами

$A_1 f(x,y) = h_1(x)\varphi^T(y) = h_1(x) B_1^{-1} F_1^T(y)$, $A_2 f(x,y) = \psi(x) h_2^T(y) = F_2(x) B_2^{-1} h_2^T(y)$,

$A_1 A_2 f(x,y) = h_1(x) B_1^{-1} F B_2^{-1} h_2^T(y)$, $h_1(x) = [h_{1,0}(x), \ldots, h_{1,N}(x)]$,

$h_2(y) = [h_{2,0}(y), \ldots, h_{2,N}(y)]$, $B_1 = \int_0^1 h_1^T(x) h_1(x) dx$, $B_2 = \int_0^1 h_2^T(y) h_2(y) dy$,

$F = \iint_G h_1^T(x) f(x,y) h_2(y) dx dy$, $F_1^T(y) = \int_0^1 h_1^T(x) f(x,y) dx$, $F_2(x) = \int_0^1 f(x,y) h_2(y) dy$.

Теоретические подходы к оценке ошибки округления

Отметим основополагающие труды, посвященные оценке ошибки округления [3-8]. В работе [9] предлагается метод оценки погрешностей округления, отличный от рассмотренных в [3-8].

Ошибка округления – это ошибка, возникающая при выполнении арифметических операций на ЭВМ с округлением результатов до фиксированного количества разрядов. Различают два режима работы ЭВМ – с фиксированной запятой и пла-

вающей запятой. Для расчетов с фиксированной запятой каждое число находится в интервале $-1 \le x \le 1$, в который исходные числа приводятся путем масштабирования. При расчетах с плавающей запятой каждое число представляется в виде $x = 2^b a$, где b – целое положительное или отрицательное число, называемое порядком, a (мантисса) – число, удовлетворяющее одному из неравенств: $-1 \le a \le -1/2$ или $1/2 \le a \le 1$ [3].

Теоретические сведения по оценке погрешности округления описаны в [5], [6].

В работе [3] комплексно анализируются важнейшие вычислительные аспекты определения ошибки математической модели и построения ее оптимальных реализаций. Комплексный подход основан на анализе трех основных характеристик вычислительных методов – точности, времени реализации, требуемой памяти ЭВМ. По данным характеристикам выполняется сравнительный анализ и оптимизация соответствующих численных методов. Отмечено, что в практике численного вычисления задач на ЭВМ применяются следующие характеристики задач, алгоритмов и ЭВМ: $E(I,X,Y)$ – полная ошибка решения E задачи P на ЭВМ с помощью алгоритма A, $T(I,X,Y)$ – время, необходимое для получения решения задачи; $M(I,X,Y)$ – необходимая память ЭВМ; fef – коэффициент технико-экономической эффективности.

В свою очередь полная абсолютная ошибка решения задачи $P(I)$ на ЭВМ $C(Y)$ с помощью вычислительного алгоритма $A(X)$ определяется так

$$\Delta(I,X,Y) = \rho\left(R, A(X,Y)\tilde{I}_p\right) = \rho(R, \tilde{R}_q).$$

$\Delta(I,X,Y) = \Delta_1 + \Delta_2' + \Delta_3$, $\Delta_2' = \rho(R_{\varepsilon,q,h}, A(X)I_p)$, где $\Delta_1 = \rho(R, R_{\varepsilon,q,h})$ – неустранимая ошибка решения задачи или ошибка за счет неточности входных данных; Δ_2' – ошибка алгоритма $A(X)$ для определения $R_{\varepsilon,q,h}$, $\Delta_3 = \rho(R_q, A(X,Y)I_p')$ – ошибка округления реализации вычислительного алгоритма $A(X)$ на ЭВМ $C(Y)$.

В дальнейших рассуждениях рассмотрим два способа нахождения обратной матрицы, определяющие подход к оценке ошибки округления при приближении функции двух переменных оператором вида (2) с использованием одномерных операторов.

В работе [7] отмечено, что значительная часть наиболее известных численных методов решения систем линейных алгебраических уравнений (далее – СЛАУ)

$$Ax = b \qquad (3)$$

основаны на разложении матрицы A на сомножители.

В зависимости от того, как связаны сомножители с матрицей A, различают две схемы построения методов. В первой схеме предполагается, что явно известны сами сомножители, на которые разложена матрица A. Пусть

$$A = B \cdot C. \qquad (4)$$

Решение системы (3) сводится к последовательному решению таких систем: $By = b$, $Cx = y$.

Во второй схеме предполагается, что найденные матрицы L, S, G, для которых выполняется соотношение

$$LAS = G. \qquad (5)$$

Тогда

$$x = Su, \quad (6)$$

где u – решение системы

$$Gu = l \quad (7)$$

с матрицею G из (5) и правой частью

$$l = Lb. \quad (8)$$

Решение системы (3) сводится теперь к вычислению вектора l согласно (8), решению системы (7) и определению искомого вектора x по формуле (6). В данной схеме матрицы L и S обычно бывают представлены в виде произведения элементарных матриц.

Разложения (4), (5) можно использовать для вычисления обратной матрицы. Для (4) следует, что

$$A^{-1} = C^{-1}B^{-1}, \quad (9)$$

а из (5) имеем

$$A^{-1} = SG^{-1}L. \quad (10)$$

Поэтому если выполнено преобразование (4) или (5), то для получения матрицы A^{-1} остается только преобразовать одну или две матрицы простого вида и осуществить умножение матриц согласно (9) или (10).

К задаче вычисления обратной матрицы можно подойти несколько иначе. Матрица A^{-1} является единственным решением матричного уравнения $AX = E$.

Обозначим через $x_1,...,x_n$ вектор-столбцы матрицы A^{-1}. Тогда x_i является решением системы линейных алгебраических уравнений

$$Ax_i = e_i, \quad (11)$$

где e_i – координатный вектор с единицей на i-м месте. Снова для решения системы (11) будут полезными разложения (4), (5).

С практической точки зрения безразлично, вычислять обратную матрицу по формуле (9), (10) или с помощью решения систем (11). Автор работы [7] предпочитает второй способ, поскольку все вопросы, связанные с решением систем, уже исследованы. При реализации этого способа может потребоваться некоторое изменение вычислительной схемы методов, вызванное необходимостью одновременного решения систем (11) со многими правыми частями.

Предположим, что выполнено преобразование (5), причем матрица G – правая треугольная, а матрица S представлена в виде произведения $U_1...U_{n-2}U_{n-1}$ матриц отражения. Последовательное решение систем (7) с правыми частями $e_1,...,e_n$ вновь позволяет разместить всю информацию о решении на месте соответствующих столбцов матрицы G. Преобразование (6) будем осуществлять последовательно, путем умножения сначала всех векторов на U_{n-1}, затем на U_{n-2} и, наконец, на U_1. При этом следует учитывать как специальный вид преобразуемых векторов, так и специальный вид самых преобразований. Элементы матрицы A^{-1} снова могут быть получены на месте матрицы A^{-1} после выполнения около $3n^3$ арифметических операций.

Для численных методов решения системы линейных алгебраических уравнений, основанных на разложении матрицы, для каждого столбца реально вычисленной матрицы $\widetilde{A}^{-1}$ и для самой матрицы A^{-1} можем записать [8]

$$\frac{\left\|\widetilde{A}^{-1} - A^{-1}\right\|_E}{\left\|A^{-1}\right\|_E} \le 2\nu_A f(n) p^{-t+1}, \qquad (12)$$

где ν_A – евклидовое число обусловленности матрицы A, $\nu_A = \left\|A^{-1}\right\|_E \|A\|_E$, $\|A\|_E = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m}(a_{i,j})^2}$.

Применение систем (11) для вычисления матрицы A^{-1} позволяет, если будет необходимо, уточнить отдельные или все ее столбцы [8].

Кроме того, провести оценку ошибки округления построения обратной матрицы можно с помощью следующих соображений [7]. Пусть $\widetilde{T}$ – реально задана или реально вычисленная по некоторому вектору b согласно предварительным условиям матрица вращения. Возьмем любой действительный вектор a. Если $\widetilde{\tau}$ и $\|a\|_E$ гораздо больше машинного нуля, где τ – количество базисных элементов, моделирующих число разрядов p-ичной системы исчисления вычислительной машины, то при умножении матрицы $\widetilde{T}$ на вектор a выполняются соотношения

$$fl(\widetilde{T}a) \equiv \widetilde{T}a + f\ , \|f\|_E \lesssim \sqrt{2}\widetilde{\tau}p^{-t+1}\|a\|_E,\ fl(\widetilde{T}a) \equiv \widetilde{T}(a+\varepsilon)\,, \|\varepsilon\|_E \lesssim \sqrt{2}p^{-t+1}\|a\|_E,$$

где $fl(\cdot)$ – результат вычислений на ЭВМ выражения в скобках, t – количество разрядов после запятой (в случае фиксированной запятой) или мантиссы (в случае плавающей запятой).

Для матрицы $\widetilde{T}$, реально вычисленной, согласно тому же условию матрица $\widetilde{T}\widetilde{T}'$ является скалярной, и при этом выполняются следующие оценки

$$\left\|\widetilde{T}\widetilde{T}' - E\right\|_2 \lesssim \frac{5}{2}p^{-t+1}, \left\|\widetilde{T} - T\right\| \lesssim \frac{5}{4}p^{-t+1},\ \widetilde{\tau} = (\widetilde{c}^2 + \widetilde{s}^2)^{1/2} = 1 + \nu\,,\ |\nu| \lesssim \frac{5}{4}p^{-t+1}.$$

Пусть по вектору b реально вычисляется матрица T и вычисляется единственная ненулевая координата вектора Tb, тогда

$$fl(Tb) \equiv \widetilde{T}(b+\varepsilon)\,,\ \|\varepsilon\|_E \lesssim \frac{\sqrt{5}}{2}p^{-t+1}\|b\|_E.$$

Приведенные оценки говорят о том, что реально вычисленная матрица вращения T с высокой степенью точности не только близка к некоторой ортогональной матрице, но даже близка к ортогональной матрице, получаемой при точных вычислениях. При этом выявляются малыми и эквивалентные возмущения преобразованных векторов.

Пусть n-мерный вектор z умножается на последовательность из N матриц вращения $T_{i_1 j_1},\ldots,T_{i_N j_N}$. Предположим, что $\widetilde{T}_{i_1 j_1},\ldots,\widetilde{T}_{i_N j_N}$ – реально заданные или реально вычисленные матрицы вращения, удовлетворяющие описанному выше условию. Тогда для любой последовательности пар индексов $i_1 j_1,\ldots,i_N j_N$ имеют место соотношения

$$fl\left(\widetilde{T}_{i_1 j_1}\ldots\widetilde{T}_{i_N j_N} z\right) \equiv \left(\widetilde{T}_{i_1 j_1}\ldots\widetilde{T}_{i_N j_N}\right)(z+\xi)\,,\ \|\xi\|_E \lesssim \sqrt{2}Np^{-t+1}\|z\|_E.$$

Отметим, если точное решение есть результат реализации некоторого алгоритма над входными данными A, а приближенно вычисленное решение можно рассматривать как результата реализации того же точного алгоритма над входными данными A_t, то отклонение A_t от A называется эквивалентным возмущением [7].

Для эквивалентного возмущения M матрицы A при разложении A на множители имеет место оценка

$$\|M\|_E \mathrel{\widetilde{\leqq}} f(n)p^{-t+1}\|A\|_E. \tag{13}$$

Тогда реально вычисленное решение $\widetilde{x}$ системы $Ax = b$ является точным решением возмущенной системы $(A+\xi)\widetilde{x} = b+\varepsilon$. При этом

$$\|\xi\|_E \le \varphi(n)p^{-t+1}\|A\|_E, \|\varepsilon\|_E \le \psi(n)p^{-t+1}\|b\|_E,$$

где $\varphi(n)+\psi(n) \mathrel{\widetilde{\leqq}} 2f(n)$, если только в пределах таких возмущений матрица остается невырожденной.

При этом выполняется неравенство

$$\frac{\|\widetilde{x}-x\|_E}{\|x\|_E} \mathrel{\widetilde{\leqq}} 2\nu_A f(n)p^{-t+1}, \tag{14}$$

которое является следствием неравенства $\|A\widetilde{x}-b\|_E \mathrel{\widetilde{\leqq}} 2f(n)p^{-t+1}\|A\|_E\|\widetilde{x}\|_E$.

Согласно (14) точность любого метода полностью определяется точностью разложения матрицы на множители.

Результаты вычислительного эксперимента

Учитывая, что в операторе (2) используются одномерные операторы, действующие на одну переменную, проведем вычислительный эксперимент и определим ошибку округления при приближении функции одной переменной $f(x)$.

На основе теоремы о виде остатка интерполяционного полинома Лагранжа степени $n-1$ функции $g(t)$ вида [10]

$$R_n g(t) = \sum_{k=0}^{n} l_{n-1,k}(t)\int_{t_k}^{t} g^{(r)}(\tau)\frac{(t_k-\tau)^{r-1}}{(r-1)!}d\tau,\ 1 \le r \le n$$

запишем минимизационное условие приближения функции $f(x)$ в виде

$$J(C) = \int_0^1\left[\sum_{k=0}^{n} C_k h(nx-k) - f(x)\right]^2 dx =$$

$$= \int_0^1\left[\sum_{k=0}^{n} C_k h(nx-k) - \sum_{k=0}^{n}\left[f\left(\frac{k}{n}\right) + \int_{\frac{k}{n}}^{x} f''(t)\frac{\left(\frac{k}{n}-t\right)}{1!}dt\right]h(nx-k)\right]^2 dx \to \min_{C_k},$$

где $h(t) = \dfrac{\left[|t+1| - 2|t| + |t-1|\right]}{2}$.

В результате для нахождения C_k, $k=\overline{0,n}$ получаем систему линейных алгебраических уравнений $\dfrac{\partial J(C)}{\partial C_p}=0$, $p=\overline{0,n}$.

Получим, положив $C_k - f_k = \varepsilon_k$, $f_k = f\left(\dfrac{k}{n}\right)$, $B_{p,k} = \int\limits_0^1 h(nx-k)h(nx-p)dx$, $0 \le k, p \le n$

$$\sum_{k=0}^{n} B_{p,k}\varepsilon_k = \sum_{k=0}^{n} F_{k,p} = \sum_{k=0}^{n}\int\limits_0^1\left[\int\limits_{\frac{k}{n}}^{x} f''(t)\left(\frac{k}{n}-t\right)dt\right]h(nx-k)h(nx-p)dx,\ p=\overline{0,n} \quad (15)$$

Из системы (15) вытекает, что, если $f''(t)\equiv 0 \Rightarrow \varepsilon_k = 0$, то $C_k = f_k$, то есть для $f(x)=b_0+b_1x$ МНК при указанной $h(t)$ является точным.

Если же $f(t)=b_0+b_1t+\dfrac{N}{2}t^2$, то есть $f''(t)=N=const$, то $F_{k,p}$, входящие в правые части имеют вид

$$F_{k,p} = -\frac{N}{2}\int\limits_0^1\left(\frac{k}{n}-x\right)^2 h(nx-k)h(nx-p)dx = \begin{cases} 0, |k-p|\ge 2, \\ -\dfrac{N}{n^3}\dfrac{3}{5!}, |k-p|=1, \\ -\dfrac{N}{n^3}\dfrac{1}{30}, |k-p|=0. \end{cases}$$

Докажем эти формулы.

При $|k-p|\ge 2$ получаем $\sup h(nx-k)\cap \sup h(nx-p)=\varnothing$, то есть $F_{k,p}=0$.

При $|k-p|=1$ получаем

$$F_{p,p+1} = -\frac{N}{2}\int\limits_{\frac{p}{n}}^{\frac{p+1}{n}}\left(\frac{p}{n}-x\right)^2(nx-p-1)(nx-p)dx =$$

$$= \frac{N}{2n^2}\int\limits_{\frac{p}{n}}^{\frac{p+1}{n}}(nx-p)^3(p+1-nx)dx = -\frac{N}{2n^3}\frac{3!\cdot 1!}{5!} = -\frac{N}{40n^3}.$$

Аналогично, $F_{p-1,p} = \dfrac{N}{40n^3}$.

При $|k-p|=0$ получаем

$$F_{p,p} = -\frac{N}{2n^2}\int\limits_{\frac{p-1}{n}}^{\frac{p+1}{n}}(p-nx)^2 h(nx-p)(nx-p)dx =$$

$$= -\frac{N}{n^2}\int\limits_{\frac{p}{n}}^{\frac{p+1}{n}}(p-nx)^2(p+1-nx)^2dx = -\frac{N}{n^2}\cdot\frac{1}{n}\cdot\frac{2!\cdot 2!}{5!} = -\frac{N}{n^3}\cdot\frac{1}{30}.$$

Тогда система (15) имеет вид

$$\sum_{k=0}^{n} B_{p,k}\varepsilon_k = \sum_{k=0}^{n} F_{p,k}\ ,\ p=\overline{0,n}\ ,$$

где, в случае $1 \le p \le n-1$, $F_{k,p} = \begin{cases} 0, \left|p-k\right| \ge 2, \\ -\dfrac{N}{n^3}\cdot\dfrac{1}{40}, \left|p-k\right| = 2, \\ -\dfrac{N}{n^3}\cdot\dfrac{1}{30}, \left|p-k\right| = 0. \end{cases}$

то есть если $p \ne 0, n$ правые части будут равняться

$$\sum_{k=0}^{n} F_{k,p} = \sum_{k=p-1}^{p+1} F_{k,p} = -\frac{N}{n^3}\cdot\frac{1}{40} - \frac{N}{n^3}\cdot\frac{1}{30} - \frac{N}{n^3}\cdot\frac{1}{40} = -\frac{N}{12n^3}.$$

Если $p=0$, $k=0$, то

$$F_{0,0} = -\frac{N}{2}\int_0^1 \left(\frac{k}{n}-x\right)^2 h(nx-k)h(nx-p)dx\Bigg|_{\substack{k=0,\\ p=0}} = -\frac{N}{2}\int_0^1 \left(\frac{0}{n}-x\right)^2 h(nx)h(nx)dx =$$

$$= -\frac{N}{2}\int_0^{\frac{1}{n}} x^2(1-nx)^2\,dx = -\frac{N}{2n^3}\frac{2!\cdot 2!}{5!} = -\frac{N}{60n^3}$$

Для случая, когда $p=0$, $k=1$, то

$$F_{0,1} = -\frac{N}{2}\int_0^1 \left(\frac{k}{n}-x\right)^2 h(nx-k)h(nx-p)dx\Bigg|_{\substack{k=1,\\ p=0}} =$$

$$= -\frac{N}{2n^2}\int_0^1 \left(nx-1\right)^2 h(nx-1)h(nx)dx = -\frac{N}{2n^2}\int_0^{\frac{1}{n}} (1-nx)^3\, nx dx = -\frac{N}{2}\frac{1}{n^3}\frac{3!\cdot 1!}{5!}$$

Тогда если $p=0$ правая часть будет иметь вид

$$\sum_{k=0}^{n} F_{k,0} = F_{0,0} + F_{1,0} = -\frac{N}{60n^3} - \frac{N}{40n^3} = -\frac{N}{24n^3}.$$

Аналогично при $p=n$ получаем

$$\sum_{k=0}^{n} F_{k,n} = F_{n-1,n} + F_{n,n} = -\frac{N}{40n^3} - \frac{N}{60n^3} = -\frac{N}{24n^3}.$$

Таким образом, система алгебраических уравнений может быть записана в виде

$$\begin{cases} \displaystyle\sum_{k=0}^{1} B_{p,k}\varepsilon_k = -\frac{N}{24n^3},\ p=0 \\ \displaystyle\sum_{k=p-1}^{p+1} B_{p,k}\varepsilon_k = -\frac{N}{12n^3},\ p=\overline{1,n-1}, \\ \displaystyle\sum_{k=n-1}^{n} B_{p,k}\varepsilon_k = -\frac{N}{24n^3},\ p=n \end{cases}$$

В условиях вычислительного эксперимента, решая СЛАУ вида (3), получаем точные значения матрицы A. Для нахождения эквивалентного возмущения M используем матрицу A_t, полученную с учетом различного количества разрядов мантиссы. Информация о норме матрицы возмущения позволяет из (13) определить $f(n) \gtrapprox \frac{\|M\|_E}{p^{-t+1}\|A\|_E}$ и далее относительную ошибку из (14).

Учитывая, что возмущение вычислительного эксперимента получено только за счет учета количества разрядов мантиссы, считаем, что величина (14) является относительной ошибкой связанной с округлением.

Тогда для функции $f(x) = x^2/2$ при $n = 10$, $p = 10$ относительная ошибка округления, возникающая при нахождении члена $z = B_1^{-1} F_1^T(y)$ в одномерных операторах формулы (2), составляет $O(10^{-9})$ при $t = 10$, $O(10^{-14})$ при $t = 15$, $O(10^{-19})$ при $t = 20$. Следует отметить, что для большего n ошибка увеличивается.

Выводы и перспективы дальнейших исследований

Выполнение различных научных исследований сегодня невозможно без применения ЭВМ для автоматизации сложных вычислительных процессов и преобразования большого количества информации. Такие преобразования в конечном счете сводятся к выполнению последовательности простейших операций, что и приводит к необходимости учитывать ошибку округления на каждом этапе. В данной работе предложен подход к оценке ошибки округления, которая возникает в ходе решения систем $B_1 u = F_1^T(y)$ и $vB_2 = F_2(x)$ путем вычисления обратной матрицы.

В дальнейших исследованиях авторы планируют рассмотреть метод оценки ошибки округления приближения функций двух переменных через одномерные операторы с использованием подходов, описанных в [9].

Список литературы

1. Литвин О.М. Наближення функціями спеціального виду функцій двох змінних, заданих дискретно або слідами на системі прямих / О.М. Литвин, О.В. Ярмош // Вісник Харківського національного університету імені В.Н. Каразіна. Серія «Математичне моделювання. Інформаційні технології. Автоматизовані системи управління»: Зб. наук. праць. – Харків, 2012. – №1015, Випуск 19. – С. 218-225.
2. Литвин О.М. Про похибку апроксимації функції двох змінних білінійними сплайнами МНК в інтегральній формі / О.М. Литвин, О.В. Ярмош // Біоніка інтелекту: наук.-техн. журнал. – 2012. - № 1(78). – С. 33-36.
3. Сергієнко І.В. Теорія оптимальних алгоритмів / І.В. Сергієнко, В.К. Задірака, О.М. Литвин. – К. : Наукова думка, 2012. – 404 с.
4. Оптимальні алгоритми обчислення інтегралів від швидкоосцилюючих функцій та їх застосування. Том 1. Алгоритми / І.В. Сергієнко, В.К. Задірака, О.М. Литвин, С.С. Мельникова, О.П. Нечуйвітер. – К. : Наукова думка, 2011. – 448 с.
5. Бабич М.Д. Округления погрешности. Энциклопедия кибернетики / М.Д. Бабич. – К. : Главная редакция Украинской Советской энциклопедии. – Т. 2. – 1974. – С. 108-109.
6. Уилкинсон Дж. Х. Алгебраическая проблема собственных значений / Дж. Х. Уилкинсон. – М.: Наука, 1970. – 564 с.
7. Воеводин В.В. Матрицы и вычисления / В.В. Воеводин, Ю.А. Кузнецов. – М.: Наука, 1984. – 320 с.
8. Воеводин В.В. Вычислительные основы линейной алгебры / В.В. Воеводин. – М.: Наука, 1977. – 303 с.
9. Бирюков А.Г. Метод оценки погрешностей округления решений задач вычислительной математики в арифметике с плавающей запятой, основанный на сравнении решений с изменяемой длиной мантиссы машинного числа / А.Г. Бирюков, А.И. Гриневич // Труды МФТИ. – 2013. – Том 5, № 2. – С. 160-174.
10. Литвин О.М. Интерполирование функций. Учеб. пособ. / О.М. Литвин. – Киев: Учеб.-метод. каб. высш. образования (УМК ВО), 1988. – 31 с.

References

1. Lytvyn O.M. Nablyzhennya funktsiyamy spetsial'noho vydu funktsiy dvokh zminnykh, zadanykh dyskretno abo slidamy na systemi pryamykh / O.M. Lytvyn, O.V. Yarmosh // Visnyk Kharkivs'koho natsional'noho universytetu imeni V.N. Karazina. Seriya «Matematychne modelyuvannya. Informatsiyni tekhnolohiyi. Avtomatyzovani systemy upravlinnya»: Zb. nauk. prats'. – Kharkiv, 2012. – #1015, Vypusk 19. – S. 218-225.
2. Lytvyn O.M. Pro pokhybku aproksymatsiyi funktsiyi dvokh zminnykh biliniynymy splaynamy MNK v intehral'niy formi / O.M. Lytvyn, O.V. Yarmosh // Bionika intelektu: nauk.-tekhn. zhurnal. – 2012. #1(78). – S. 33-36.
3. Serhiyenko I.V. Teoriya optymal'nykh alhorytmiv / I.V. Serhiyenko, V.K. Zadiraka, O.M. Lytvyn. – K.: Naukova dumka, 2012. – 404 s.
4. Optymal'ni alhorytmy obchyslennya intehraliv vid shvydkoostsylyuyuchykh funktsiy ta yikh zastosuvannya. Tom 1. Alhorytmy / I.V. Serhiyenko, V.K. Zadiraka, O.M. Lytvyn, S.S. Mel'nykova, O.P. Nechuyviter. – K.: Naukova dumka, 2011. – 448 s.
5. Babich M.D. Okruglenija pogreshnosti. Jenciklopedija kibernetiki / M.D. Babich. – K.: Glavnaja redakcija Ukrainskoj Sovetskoj jenciklopedii. – T. 2. – 1974. – S. 108-109.
6. Uilkinson Dzh. H. Algebraicheskaja problema sobstvennyh znachenij / Dzh. H. Uilkinson. – M.: Nauka, 1970. – 564 s.
7. Voevodin V.V. Matricy i vychislenija / V.V. Voevodin, Ju.A. Kuznecov. – M.: Nauka, 1984. – 320 s.
8. Voevodin V.V. Vychislitel'nye osnovy linejnoj algebry / V.V. Voevodin. – M.: Nauka, 1977. – 303 s.
9. Birjukov A.G. Metod ocenki pogreshnostej okruglenija reshenij zadach vychislitel'noj matematiki v arifmetike s plavajushhej zapjatoj, osnovannyj na sravnenii reshenij s izmenjaemoj dlinoj mantissy mashinnogo chisla / A.G. Birjukov, A.I. Grinevich // Trudy MFTI. – 2013. – Tom 5, № 2. – S. 160-174.
10. Litvin O.M. Interpolirovanie funkcij. Ucheb. posob. / O.M. Litvin. – Kiev: Ucheb.-metod. kab. vyssh. obrazovanija (UMK VO), 1988. – 31 s.

RESUME

O.M. Lytvyn, O.V. Iarmosh

Estimation of Rounding Errors of Two Variables Functions Approximation of One-Dimensional Operators

The article discusses approaches to assessing rounding error when applying the method of approximation of two variables functions defined on a system of tracks perpendicular straight lines when approximation operator is given in the form of a Boolean operator sum through dimensional operators acting on the one variable function [1].

On the assumption that the rounding error occurs in the calculation of inverse matrices, considered two ways of finding the inverse matrix: according to the first method of solving linear algebraic equations systems $Ax = b$, that is search A^{-1} based on the decomposition of the matrix A into factors, the second method is based on the assumption that the matrix A^{-1} is the only solution of the matrix equation $AX = E$.

In each case, the inequality given to estimate the relative error of solving systems $Ax = b$ by inverse matrix according to [7], [8].

The results of numerical experiments have shown that the one variable function $f(x) = x^2/2$ when $n = 10$, $p = 10$ relative rounding error is $O(10^{-9})$ at $t = 10$, $O(10^{-14})$ at $t = 15$, $O(10^{-19})$ at $t = 20$ (t – the number of digits after the decimal point (in the case of fixed-point) or the mantissa (in the case of floating-point) p-ary system of any computer calculation). In addition, for more n error increases.

Статья поступила в редакцию 20.12.2013.